



United States
Environmental Protection
Agency

Office of Water
Washington, DC 20460

EPA-821-B-01-014
December 2001

Statistical Analysis of Abandoned Mine Drainage in the Assessment of Pollution Load ("The Griffiths Report")



**Statistical Analysis of
Abandoned Mine Drainage in the
Assessment of Pollution Load**

(“The Griffiths Report”)

Prepared for:

**U.S. Environmental Protection Agency
Office of Water
Office of Science and Technology
Engineering and Analysis Division**

Prepared by:

**John C. Griffiths
Roger J. Hornberger, Pennsylvania DEP
Ken Miller, DynCorp I & ET
Michael W. Smith, Pennsylvania DEP**

December 2001

DEDICATION

John C. Griffiths was one of the early leaders in the use of statistics in the geological sciences. As an attest to his world class stature, he was the first recipient of the William C. Krumbein Award by the International Association of Mathematical Geology in 1977, named after one of his contemporaries. Griffiths, Krumbein, Felix Chayes and a few others introduced geologists and geological students to statistical methods in sampling, experimental design, petrology, mineralogy, sedimentology, stratigraphy and other aspects of the geosciences throughout the 1950's, 1960's and 1970's as documented in approximately 100 scientific papers and several text books.

John Griffiths was born on February 29, 1912 in Wales. He earned 3 degrees from the University of Wales including a PhD in 1937 in glacial geology and petrography, a Diploma of Imperial College at the Royal College of Science in London, and a second PhD from the University of London in 1940. He was employed as a research petrographer on oil well drilling projects from 1940 to 1947 in Trinidad, where he was married on July 26, 1941. He was a professor in geosciences at the Pennsylvania State University from 1947 to his retirement in 1977, and thereafter was a Professor Emeritus until his death on June 2, 1992 in State College, PA. The day before his death at age 80, he was conducting research in the Earth and Mineral Sciences Library at Penn State. During his many years as a professor, he served as the Head of the Department of Mineralogy (and Geochemistry) from 1955 to 1966, and as the Director of Planning Research for the entire University from 1969 to 1971.

Dr. Griffiths was an excellent teacher who instilled scientific rigor and an appreciation for proper sampling and the use of statistics in the minds of many students. While at Pennsylvania State University, he taught univariate statistics, bivariate statistics, and multivariate statistics; periodically, he also taught a course in time series analysis. New graduate students, relying upon the foundation of their undergraduate studies, would be confronted by this feisty Welshman, armed with more than 20 years of data on a local stratified gravel deposit from previous classes, saying things like "You call yourselves geologists; you can't even tell me how many layers there are in this gravel deposit." Students soon learned that Dr. Griffiths was challenging them to use statistical analysis as a guide to the unknown in a scientific method for solving problems in the geosciences.

J.C. Griffiths approached teaching, research and much of life in general, with a blend of humor, history, and lessons learned from other sciences, observations from current events, and a strong foundation of scientific rigor and ethics. With the advances in computer science in the 1950's and 1960's, Griffiths expanded his areas of interest into related fields of computer modeling, operations research and cybernetics. In the 1960's and 1970's, he proposed drilling the entire United States on a 20-mile grid spacing, wherein approximately 7500 drill holes each 10,000 to 15,000 feet deep would almost certainly result in the discovery of billions of dollars worth of oil, gold, uranium, zinc, copper and other valuable minerals overlooked by conventional "hit-and-miss" type of exploration. In the early days of research on the correlation between cigarette smoking and the incidence of lung cancer, Griffiths was requested to meet with a famous statistical researcher for dinner the evening before his cancer research speech at the University. Griffiths was a smoker at that time, and his recollection of the evening was, "I

took one look at that man's statistics and I knew that I had 2 choices: I either had to give up cigarettes or give up statistics."

Following his retirement from the full time faculty in 1977, J. C. Griffiths worked with the U. S. Geological Survey in Reston VA and continued his research with graduate students on quantifying the geology of the world by country for mineral resource assessment purposes. He served as a consultant to DER (now DEP) and EPA from 1984 to 1988 on a cooperative project to support development of Pennsylvania's Coal Remining regulatory package.

Beyond his many professional accomplishments, John C. Griffiths was a great person. This document was prepared in his honor and with great respect for his accomplishments as a geostatistician, a teacher, and a major contributor to our understanding of sedimentary and geochemical processes.

DISCLAIMER

The statements in this document are intended solely as guidance. This document is not intended, nor can it be relied upon, to create any rights enforceable by any party in litigation with the United States. EPA may decide to follow the guidance provided in this document, or to act at variance with the guidance, based on its analysis of the specific facts presented. This guidance is being issued in connection with amendments to the Coal Mining Point Source Category.

TABLE OF CONTENTS

LIST OF FIGURES	i
LIST OF TABLES	vii
Chapter 1 Introduction	1-1
Chapter 2 Statistical Analysis of Mine Drainage Data	2-1
Chapter 3 Mine Drainage Analysis Algorithm	3-1
Chapter 4 Analysis of Data from the Arnot Site	4-1
Chapter 5 Analysis of Data from the Clarion Site	5-1
Chapter 6 Analysis of Data from the Ernest Site	6-1
Chapter 7 Analysis of Data from the Fisher Site	7-1
Chapter 8 Analysis of Data from the Markson Site	8-1
Chapter 9 Statistical Summary and Review of Quality Control Limits	9-1
APPENDIX A: Hamilton Discharge Data	A-1
APPENDIX B: Arnot Discharge Data	B-1
APPENDIX C: Clarion Discharge Data	C-1
APPENDIX D: Ernest Discharge Data	D-1
APPENDIX E: Fisher Discharge Data	E-1
APPENDIX F: Markson Discharge Data	F-1
REFERENCES	R-1

LIST OF FIGURES

Page

Chapter 1.0 Introduction

Figure 1.1: Map of Pennsylvania Counties and Mine Sites	1-3
---	-----

Chapter 2.0 Statistical Analysis of Mine Drainage Data

Figure 2.1: Example of Acid Load Variation Before, During, and After Remediation	2-1
Figure 2.2: Example of Normal Distribution	2-5
Figure 2.3: Example of a Quality Control Graph	2-6
Figure 2.4: Stem-and-leaf of Discharge	2-7
Figure 2.5: Net Alkalinity Boxplot for Fisher Mine Site Discharge	2-9

Chapter 3.0 Mine Drainage Data Analysis Algorithm

Figure 3.1: Algorithm for Analysis of Mine Drainage Discharge Data	3-3
--	-----

Chapter 4.0 Analysis of Data from the Arnot Site

Figure 4.1: Map of Arnot Site	4-2
Figure 4.2: Log Flow vs. Time (Arnot 001, 003, and 004)	4-3
Figure 4.3: Stem-and-leaf of Sulfate	4-9
Figure 4.4: Stem-and-leaf of Discharge	4-9
Figure 4.5: Stem-and-leaf of pH.	4-10
Figure 4.6: Stem-and-leaf of Sulfate	4-10
Figure 4.7: Stem-and-leaf of Acidity	4-11
Figure 4.8: Plot of Manganese vs. Log Flow	4-11
Figure 4.9: Plot of Acidity vs. Flow	4-12
Figure 4.10: Plot of Manganese vs. Flow	4-12
Figure 4.11a: Plot of Discharge vs. Time	4-14
Figure 4.11b: Plot of pH vs. Time	4-14
Figure 4.11c: Plot of Acidity vs. Time	4-15
Figure 4.11d: Plot of Total Iron vs. Time	4-15
Figure 4.11e: Plot of Sulfate vs. Time	4-15
Figure 4.11f: Plot of Aluminum vs. Time	4-16
Figure 4.12a: Autocorrelation Function of Discharge	4-19
Figure 4.12b: Autocorrelation Function of Calcium	4-19
Figure 4.12c: Autocorrelation Function of Aluminum	4-20

Figure 4.12d: Autocorrelation Function of Total Iron	4-20
Figure 4.13a: Autocorrelation Function of pH.	4-21
Figure 4.13b: Partial Autocorrelation Function of pH.	4-22
Figure 4.13c: Autocorrelation Function of Log Discharge	4-22
Figure 4.13d: Partial Autocorrelation Function of Log Discharge	4-23
Figure 4.13e: Autocorrelation Function of Ferric Iron	4-23
Figure 4.13f: Partial Autocorrelation Function of Ferric Iron	4-24

Chapter 5.0 Analysis of Data from the Clarion Site

Figure 5.1: Map of Clarion Mine Site	5-2
Figure 5.2: Stem-and-leaf of Sulfate	5-4
Figure 5.3a: Stem-and-leaf of Discharge	5-5
Figure 5.3b: Stem-and-Leaf of Log Discharge	5-5
Figure 5.4a: Stem-and-leaf of Acid	5-6
Figure 5.4b: Stem-and-Leaf of Log Acid	5-6
Figure 5.5a: Cross Correlation Function for pH and Discharge	5-8
Figure 5.5b: Cross Correlation Function for Acidity and Discharge	5-9
Figure 5.5c: Cross Correlation between Sulfate and Discharge	5-9
Figure 5.6a: Plot of pH vs. Acidity	5-10
Figure 5.6b: Plot of Acidity vs. Sulfate	5-10
Figure 5.6c: Plot of Iron vs. Sulfate	5-11
Figure 5.7a: Plot of pH vs. Time	5-12
Figure 5.7b: Plot of Acidity vs. Time	5-13
Figure 5.7c: Plot of Iron vs. Time	5-13
Figure 5.7d: Plot of Sulfate vs. Time	5-13
Figure 5.8a: Autocorrelation Function of pH.	5-14
Figure 5.8b: Autocorrelation Function of Acid	5-15
Figure 5.8c: Autocorrelation Function of Total Iron	5-15
Figure 5.8d: Autocorrelation Function of Sulfate	5-16
Figure 5.9: Projections of Sulfate Data	5-18

Chapter 6.0 Analysis of Data from the Ernest Site

Figure 6.0: Map of Ernest Site	6-2
Figure 6.1a: Histogram of pH (n = 174)	6-5
Figure 6.1b: Histogram of Acid Load (n = 174)	6-5
Figure 6.1c: Histogram of Log Acid Load (n=174)	6-5
Figure 6.1d: Histogram of Total Iron (n=174)	6-6
Figure 6.1e: Histogram of Log Total Iron (n=174)	6-6
Figure 6.1f: Histogram of SO ₄ (n=174)	6-6
Figure 6.1g: Histogram of Log SO ₄ (n=174)	6-7
Figure 6.2a: Cross Correlation Function of pH vs. Flow	6-9

Figure 6.2b:	Cross Correlation Function of pH vs. Log Acid.	6-9
Figure 6.2c:	Cross Correlation Function of pH vs. Log Acid Load	6-10
Figure 6.2d:	Cross Correlation Function of pH vs. Log Iron	6-10
Figure 6.3a:	Plot of pH vs. Time (Days)	6-11
Figure 6.3b:	Plot of Log Flow vs. Time (Days)	6-11
Figure 6.3c:	Plot of Log Acid vs. Time (Days)	6-12
Figure 6.4:	Plot of Acid vs. Acid Load	6-12
Figure 6.5a:	Plot of Iron Load vs. Acid Loading	6-13
Figure 6.5b:	Plot of Acid Load vs. Sulfate Loading	6-13
Figure 6.5c:	Plot of Iron vs. Sulfate Loading	6-14
Figure 6.6a:	Bivariate Plot of Log Acidity vs. Log Flow	6-14
Figure 6.6b:	Bivariate Plot of Log Sulfate vs. Log Flow	6-15
Figure 6.6c:	Bivariate Plot of Log Acidity vs. Log Acid Load	6-15
Figure 6.6d:	Bivariate Plot of Log Sulfate vs. Log Acid	6-16
Figure 6.7a:	Time Series Plot of pH.	6-17
Figure 6.7b:	Time Series Plot of Flow	6-17
Figure 6.7c:	Time Series Plot of Acidity	6-18
Figure 6.7d:	Time Series Plot of Iron Load	6-18
Figure 6.7e:	Time Series Plot of Acid Load	6-19
Figure 6.7f:	Time Series Plot of Sulfate Load	6-19
Figure 6.8a:	Autocorrelation Function of pH.	6-22
Figure 6.8b:	Autocorrelation Function of Iron.	6-22
Figure 6.8c:	Autocorrelation Function of Flow	6-23
Figure 6.8d:	Partial Autocorrelation Function of Flow	6-23
Figure 6.8e:	Autocorrelation Function of Acidity	6-24
Figure 6.8f:	Partial Autocorrelation Function of Acid	6-24
Figure 6.8g:	Autocorrelation Function of Acid Load	6-25
Figure 6.8h:	Partial Autocorrelation Function of Acid Load	6-25

Chapter 7.0 Analysis of Data from the Fisher Site

Figure 7.0:	Map of Fisher Site	7-2
Figure 7.1a:	Histogram of Log Flow	7-6
Figure 7.1b:	Histogram of Log Acid	7-6
Figure 7.1c:	Histogram of Log Iron	7-6
Figure 7.1d:	Histogram of Log Manganese	7-7
Figure 7.1e:	Histogram of Log Aluminum	7-7
Figure 7.2a:	Log Flow vs. Time (Days)	7-9
Figure 7.2b:	Log Acidity vs. Time (Days)	7-10
Figure 7.2c:	Log Iron vs. Time (Days)	7-11
Figure 7.2d:	Log Manganese vs. Time (Days)	7-12
Figure 7.2e:	Log Aluminum vs. Time (Days)	7-13
Figure 7.3:	Plot of Log Iron vs. Acidity	7-17
Figure 7.4a:	Collection Dates vs. Observation Number (First Differences)	7-19

Figure 7.4b:	Plot of Log Flow vs. Time	7-19
Figure 7.4c:	Plot of Log Acidity vs. Time	7-20
Figure 7.4d:	Plot of Log Sulfate vs. Time	7-20
Figure 7.4e:	Plot of Log Iron vs. Time	7-20
Figure 7.4f:	Plot of Log Manganese vs. Time	7-21
Figure 7.4g:	Plot of Log Aluminum vs. Time	7-21
Figure 7.5a:	Autocorrelation Function of Days	7-23
Figure 7.5b:	Partial Autocorrelation Function of Days	7-23
Figure 7.5c:	Autocorrelation Function of Flow	7-23
Figure 7.5d:	Partial Autocorrelation Function of Flow	7-24
Figure 7.5e:	Autocorrelation Function of Acid	7-24
Figure 7.5f:	Partial Autocorrelation Function of Acid	7-24
Figure 7.5g:	Autocorrelation Function of Sulfate	7-25
Figure 7.5h:	Partial Autocorrelation Function of Sulfate	7-25
Figure 7.5i:	Autocorrelation Function of Iron	7-25
Figure 7.5j:	Partial Autocorrelation Function of Iron	7-26
Figure 7.5k:	Autocorrelation Function of Manganese	7-26
Figure 7.5l:	Partial Autocorrelation Function of Manganese	7-26
Figure 7.5m:	Autocorrelation Function of Aluminum	7-27
Figure 7.5n:	Partial Autocorrelation Function of Aluminum	7-27

Chapter 8.0 Analysis of Data from the Markson Site

Figure 8.0:	Map of Markson Site	8-2
Figure 8.1a:	Histogram of Interval (n=106)	8-5
Figure 8.1b:	Histogram of Log Flow (n=107)	8-6
Figure 8.1c:	Histogram of Log Acidity (n=252)	8-6
Figure 8.1d:	Histogram of Log Acidity (n=107)	8-6
Figure 8.1e:	Histogram of Iron (n=253)	8-7
Figure 8.1f:	Histogram of Manganese (n=249)	8-7
Figure 8.1g:	Histogram of Log Aluminum (n=246)	8-7
Figure 8.1h:	Histogram of Sulfate (n=253)	8-8
Figure 8.1i:	Histogram of Log Ferrous Iron (n=241)	8-8
Figure 8.2a:	Plot of Sulfate vs. Log Flow	8-10
Figure 8.2b:	Plot of Iron vs. Log Flow	8-10
Figure 8.2c:	Plot of Manganese vs. Log Aluminum	8-11
Figure 8.2d:	Plot of Log Aluminum vs. Sulfate	8-11
Figure 8.2e:	Plot of Manganese vs. Log Flow	8-11
Figure 8.2f:	Plot of Log Acidity vs. Log Flow	8-12
Figure 8.2g:	Plot of Log Acid vs. Iron	8-12
Figure 8.3a:	Time Series Plot of Flow	8-14
Figure 8.3b:	Time Series Plot of Manganese	8-15
Figure 8.3c:	Time Series Plot of Sulfate	8-16

Figure 8.3d:	Time Series Plot of Iron	8-17
Figure 8.3e:	Time Series Plot of Aluminum.	8-18
Figure 8.3f:	Time Series Plot of Acidity.	8-19
Figure 8.4a:	Autocorrelation Function of Flow	8-22
Figure 8.4b:	Autocorrelation Function of Iron	8-23
Figure 8.4c:	Partial Autocorrelation Function of Iron	8-23
Figure 8.4d:	Autocorrelation Function of Aluminum.	8-24
Figure 8.4e:	Partial Autocorrelation Function of Aluminum.	8-24
Figure 8.4f:	Autocorrelation Function of Intrvals	8-25
Figure 8.4g:	Autocorrelation Function of pH.	8-25
Figure 8.4h:	Partial Autorcorrelation Function of pH.	8-26
Figure 8.4i:	Autocorrelation Function of Manganese	8-26
Figure 8.4j:	Partial Autocorrelation Function of Manganese	8-27
Figure 8.4k:	Autocorrelation Function of Sulfate	8-27
Figure 8.4l:	Partial Autocorrelation Function of Sulfate	8-28

Chapter 9.0 Statistical Summary and Review of Quality Control Limits

Figure 9.1:	Example graph Log Discharge versus Days	9-5
Figure 9.2:	Flow Chart for Box-Jenkins Time Series Analysis	9-9

APPENDIX A: Hamilton Discharge Data

Figure A.1:	Map of Hamilton Site	A-2
-------------	----------------------------	-----

LIST OF TABLES

Chapter 1.0 Introduction

Chapter 2.0 Statistical Analysis of Mine Drainage Data

Chapter 3.0 Mine Drainage Data Analysis Algorithm

Table 3.1:	Summary Statistics for S3CLAR (N=96)	3-5
------------	--	-----

Chapter 4.0 Analysis of Data from the Arnot Site

Table 4.1a:	Summary Statistics for Arnot 001 Data	4-5
Table 4.1b:	Summary Statistics for Arnot 003 Data	4-6
Table 4.1c:	Summary Statistics of Arnot 004 Data	4-7
Table 4.2:	Coefficient of Variation (%)	4-8
Table 4.3:	Comparison of Confidence Belts Around Mean and Median	4-17
Table 4.4:	Observations Falling Beyond Confidence Limits of 2 Standard Deviations Around the Mean Beyond the 1.58* H-Spread	4-17
Table 4.5:	Comparison of Total Iron and Ferrous Iron	4-21

Chapter 5.0 Analysis of Data from the Clarion Site

Table 5.1:	Summary Statistics for S3CLAR (n=96)	5-3
Table 5.2:	Summary Statistics for S3CLAR Adjusted Data Deck (n=79)	5-7
Table 5.3:	Summary of Time Series Models for pH, Clarion Mine	5-17
Table 5.4:	Summary Statistics for Time Series Models of SO ₄ from Clarion Site	5-17
Table 5.5:	Summary of Time Series Models for Ferrous Iron Clarion Site	5-19
Table 5.6:	Comparison of Statistics Used to Calculate the QC Limits	5-20
Table 5.7:	Comparison of QC Limits (Spreads) around Mean and Median	5-21
Table 5.8:	Comparison of QC Limits Around the Median Using Various Forms of Spread	5-22

Chapter 6.0 Analysis of Data from the Ernest Site

Table 6.1:	Summary Statistics of Data (n=174)	6-3
Table 6.2:	Summary Statistics of Log-transformed Data (n=174)	6-4
Table 6.3:	Correlation Coefficients for 9 Parameters (n=174)	6-8
Table 6.4:	Base Data for Calculation of Quality Control Limits of Ernest Data	6-20

Table 6.5:	Two Measures of Quality Control	6-20
------------	---------------------------------------	------

Chapter 7.0 Analysis of Data from the Fisher Site

Table 7.1:	Summary Statistics for 79 Log Transformed Observations	7-3
Table 7.2a:	Summary statistics for 57 Log Transformed Observations (pre-remining)	7-4
Table 7.2b:	Summary statistics for 19 Log Transformed Observations (during remining)	7-5
Table 7.3:	Correlations of Zero Order Among the Seven Variables	7-15
Table 7.4:	Summary of the Important Cross-Correlations (CCF) Among Seven Variables	7-16
Table 7.5:	Equations of Models Fitted to Variables from the Fisher Deep Mine	7-28

Chapter 8.0 Analysis of Data from the Markson Site

Table 8.1:	Summary Statistics (n=107)	8-3
Table 8.2:	Summary Statistics (n = 253)	8-4
Table 8.3:	Correlations Among Variables (n = 107)	8-9
Table 8.4:	Correlations Among Variables (n = 253)	8-9
Table 8.5:	Cross-correlations Among the Variables	8-13
Table 8.6:	Quality Control Limits $\bar{X} \pm 2\sigma$	8-20
Table 8.7:	Quality Control Limits: $\bar{M}_d \pm [1.96(Q_1-Q_3)/1.35\sqrt{n}]$	8-20
Table 8.8:	Tests of the Different Models for Each Parameter	8-29
Table 8.9:	Model Equations for the Variables (see Table 8.8)	8-31

Chapter 9.0 Statistical Summary and Review of Quality Control Limits

Table 9.1:	Guidance and Protocols for Water Sample Collection	9-1
Table 9.2:	Acf of the Residuals from Fitting an MA Model to the Original Observations After Taking a First Difference: SO_4	9-12
Table 9.3:	Pacf of the Residuals from Fitting an MA Model to the Original Observations After Taking a First Difference: SO_4	9-12
Table 9.4:	Arrangement of Significant Deviations of the Residuals (> 2 Standard Error = 66.02)	9-12

Appendix A

Table A.1:	Alternate Models Fitted to Log Data	A-4
------------	---	-----